A Novel Joint Framework for Multiple Chinese Events Extraction

Nuo Xu[1,2](✉), Haihua Xie[2], and Dongyan Zhao[1]

[1] Wangxuan Institute of Computer Technology, Peking University, Beijing 100080, China
{xunuo2019,zhaodongyan}@pku.edu.cn

[2] State Key Laboratory of Digital Publishing Technology, Peking University Founder Group Co., Ltd., Beijing 100871, China
xiehh@founder.com

Abstract. Event extraction is an essential yet challenging task in information extraction. Previous approaches have paid little attention to the problem of roles overlap which is a common phenomenon in practice. To solve this problem, this paper defines event relation triple to explicitly represent relations among triggers, arguments and roles which are incorporated into the model to learn their inter-dependencies. A novel joint framework for multiple Chinese events extraction is proposed which jointly performs predictions for event triggers and arguments based on shared feature representations from pre-trained language model. Experimental comparison with state-of-the-art baselines on ACE 2005 dataset shows the superiority of the proposed method in both trigger classification and argument classification.

Keywords: Chinese multiple event extraction · Pre-trained language models · Roles overlap problem · Event relation triple

1 Introduction

Event extraction (EE) is of utility and challenge task in natural language processing (NLP). It aims to identify event triggers of specified types and their arguments in text. As defined in Automatic Content Extraction (ACE) program, the event extraction task is divided into two subtasks, i.e., trigger extraction (identifying and classifying event triggers) and argument extraction (identifying arguments and labeling their roles).

Chinese event extraction is a more difficult task because of language specific issue in Chinese [1]. Since Chinese does not have delimiters between words, segmentation is usually a necessary step for further processing, leading to word-trigger mismatch problem [2]. The approaches based on word-wise classification paradigm commonly suffer from this. It is hard to extract accurately when

Supported by National Key Research and Development Program (No. 2019YFB1406302), China Postdoctoral Science Foundation (NO. 2020M670057) and Beijing Postdoctoral Research Foundation (No. ZZ2019-92).

M. Sun et al. (Eds.): CCL 2020, LNAI 12522, pp. 174–183, 2020.
https://doi.org/10.1007/978-3-030-63031-7_13

a trigger is part of a word or cross multiple words. To avoid this issue, we formulate Chinese event extraction as a character-based classification task. In addition, another interesting issue in event extraction which is rarely followed requires more efforts. It is the roles overlap problem that we concern in this paper, including the problems of either roles sharing the same argument or arguments overlapping on some words. There are multiple events existing in the one sentence, which commonly causes the roles overlap problem and is easy to overlook [3]. Figure 1(a) shows example of roles sharing the same argument in ACE 2005 dataset. “控” (accuse) triggers a Charge-Indict event and “杀害” (kill) triggers an Attack event, while argument “他们” (them) plays the role “Defendant” as well as the role “Attacker” at the same time. Figure 1(b) shows example of arguments overlapping on some words in ACE 2005 dataset. “来往” (traveled between) triggers a Transport event, while argument “中国” (China) plays not only the role “Origin” but “Destination” and argument “来往于中国和澳大利亚之间的乘客” (passengers who traveled between China and Australia) plays the role “Artifact”. We observe that the above two arguments overlap on word “中国” (China), which is more challenging for traditional methods to simultaneously identify these two arguments, especially for those being long noun phrases. Research shows that there exist about 10% events in ACE 2005 dataset [4] having the roles overlap problem [3]. Moreover, the results of event extraction could affect the effectiveness of many other NLP tasks, such as the construction of knowledge graph. Therefore, the roles overlap problem is of great importance and needs to be seriously addressed.

It is thus appealing to design a single architecture to solve the problem. In this paper, we propose a single framework to jointly extract triggers and arguments. Inspired by the effectiveness of pre-trained language models, we adopt bidirectional encoder representation from transformer (BERT) as the encoder to obtain the shared feature representations. Specifically, the relations among triggers (t), arguments (a) and roles (r) are defined as event relation triples $<t, r, a>$ where r represents the dependencies of a on t in the event triggered by t. The event sentence of Fig. 1(b) could be represented by event relation triples as <来往, , Origin, 中国>, <来往, Destination, 中国>, <来往, Origin, 澳大利亚>, <来往, Destination, 澳大利亚>, <来往, Artifact, 来往于中国和澳大利亚之间的乘客> . As is seen, event relation triples could explicitly describe relations among the three items. The task of argument classification is converted to relation extraction. Specially, to extract multiple events and relation triples, we utilize multiple sets of binary classifiers to determine the spans (each span includes a start and an end). By this approach, not only roles overlap problem but also word-trigger mismatch and word boundary problems in Chinese language are solved. Our framework avoids human involvements and elaborate engineering features in event extraction, but yields better performance over prior works.

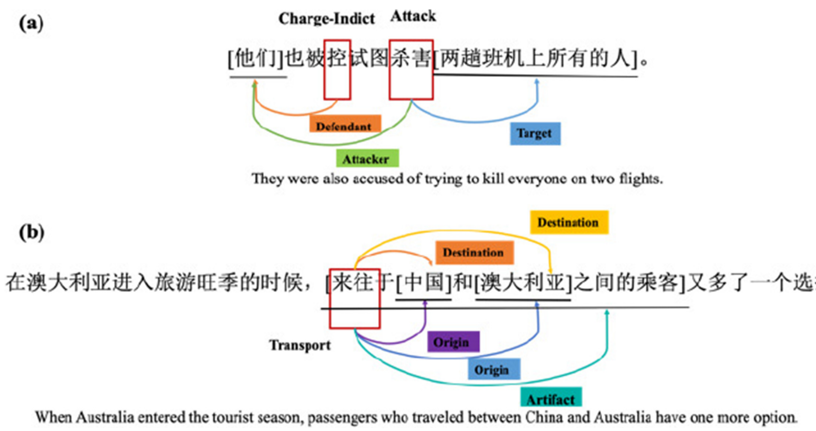

Fig. 1. Examples of roles overlap problem

2 Related Work

EE is an important task which has attracted many attentions. There are two main paradigms for EE: a) the joint approach that predicts event triggers and arguments jointly [5,6], and b) the pipelined approach that first identifies trigger and then identifies arguments in separate stages [7]. The advantages of such a joint system are twofold: (1) mitigating the error propagation from the upstream component (trigger extraction) to the downstream classifier (argument extraction), and (2) benefiting from the inter-dependencies among event triggers and argument roles [8]. Traditional methods that rely heavily on hand-craft features are hard to transfer among languages and annotation standards [9–11]. The neural network based methods that are able to learn features automatically [12,13] have achieved significant progress. Most of them have followed the pipelined approach. Some improvements have been made by jointly predicting triggers and arguments [6–8] and introducing more complicated architectures to capture larger scale of contexts. These methods have achieved promising results in EE.

Although roles overlap problem has been put forward [3,5,6], there are only few works in the literature to study this. He and Duan construct a multi-task learning with CRF enhanced model to jointly learn sub-events [5]. However, their method relies on hand-crafted features and patterns, which makes them difficult to be integrated into recent neural models. Yang et al. adopt a two-stage event extraction by adding multiple sets of binary classifiers to solve roles overlap problem which suffers from error propagation [3]. It does not employ shared feature representations as we do in this work.

In recent years, pre-trained language models are successful in capturing words semantic information dynamically by considering their context. McCann et al.(2017) pre-trained a deep LSTM encoder for machine translation (MT) to contextualize word vectors [14]. ELMo (Embeddings from Language Models) improved 6 challenging NLP problems by learning the internal states of the stacked bidirectional LSTM (Long Short-Term Memory) [15]. Open AI GPT (Generative Pre-Training) improved the state-of-the-art in 9 of 12 tasks [16]. BERT obtained new state-of-the-art results on 11 NLP tasks [17].

3 Extraction Model

This section describes our approach that is designed to extract events. We now define the scope of our work. The task of argument extraction is defined as automatically extracting event relation triples defined. In our model, instead of treating entity mentions as being provided by human annotators, only event label types and argument role types are utilized as training data for both trigger and argument extraction.

We propose a pre-trained language model based joint multiple Chinese event extractor (JMCEE). Let $s = \{c_1, c_2, \&, c_n\}$ be annotated sentence s with n as the number of characters and c_i as the ith character. Given the set of event relation triples $E = \{<t, r, a>\}$ in s, the goal of our framework is to perform the task of trigger extraction T and argument extraction A jointly:

$$P(A, T|s) = P(A|T, s) \times P(T|s) = \prod_{(r,a) \in E|t} p((r, a)|t, s) \prod_{t \in E} p(l, t|s) \quad (1)$$

Here $(r, a) \in E|t$ denotes an argument and role pair (r, a) in the event triples E triggered by t and l denotes the event label type. Based on Eq. (1), we first predict all possible triggers and their label types in a sentence; then for each trigger, we integrate information of predicted trigger word to extract event relation triple $<t, r, a>$ by simultaneously predicting all possible roles and arguments, as illustrated in Fig. 2. We employ a pre-trained BERT encoder to learn the representation for each character in one sentence, then feed it into downstream modules. Token [CLS] and [SEP] are placed at the start and end of the sentence. Multiple sets of binary classifiers are added on the top of the BERT encoder to implement predictions for multiple events and relation triples. For trigger extraction, we need to predict the start and end of event type l for $c_i \in s$ (l could be "Other" type to indicate that there is no word triggering any event) with each set of binary classifiers severing for an event type to determine the starts and ends of all triggers. For argument extraction, we need to extract event relation triple $<t, r, a>$ by predicting the start and end of role type r for c_i in sentence s based on predicted triggers (r is set to "Other" if there is no word triggering any event as well) with each set of binary classifiers severing for a role to determine the starts and ends of all arguments that play it. The roles overlap problem could be solved since the prediction could belong to different arguments and roles. Besides, our JMCEE enables to identify those arguments being long noun phrases like "来往于中国和澳大利亚之间的乘客" (passengers who traveled between China and Australia), which tackles the word boundary problem often encountered in Chinese. Compared with sentence-level sequential modeling methods, our approach also avoids suffering low ef ciency in capturing very long-range dependencies in previous works [18].

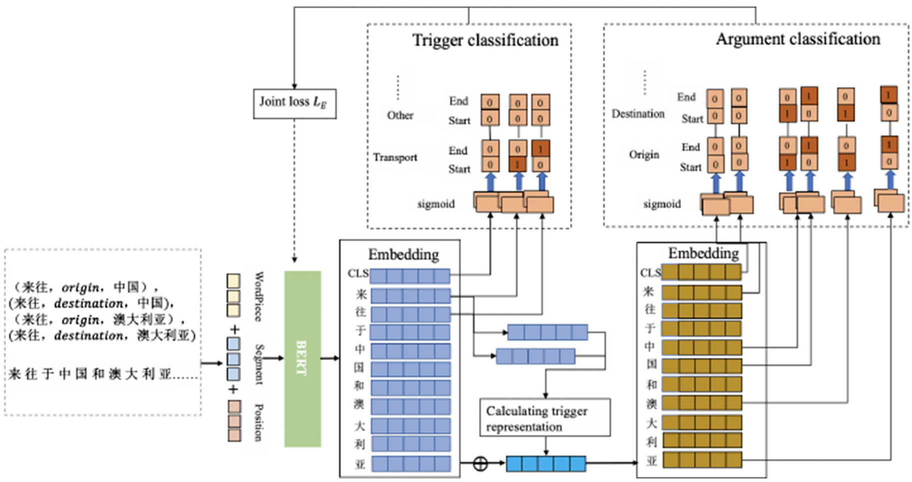

Fig. 2. The framework of JMCEE, including the trigger extract component and the argument extract component. The extraction procedure of the event instance is shown.

3.1 Trigger Extraction

Trigger extraction aims to predict whether a token is a start or an end of a trigger for type label l. A token c_i is predicted as the start of a trigger with probability for type label l through feeding it into a fully-connected layer with sigmoid activation function:

$$P_{Ts}^l(c_i) = \sigma(W_{Ts}^l \beta(c_i) + b_{Ts}^l) \tag{2}$$

while as the end with probability:

$$P_{Te}^l(c_i) = \sigma(W_{Te}^l \beta(c_i) + b_{Te}^l) \tag{3}$$

where we utilize subscript "s" to denote "start" and subscript "e" to denote "end". W_{Ts} and b_{Ts} are respectively the trainable weights and bias of binary classifier that targets to detect starts of triggers' labels, while W_{Te} and b_{Te} are respectively the trainable weights and bias of another binary classifier that targets to detect ends of triggers' labels. β is the BERT embedding. Set thresholds of detecting starts and ends as $\delta^l = \{\delta_s^l, \delta_e^l\}$, δ_s^l and δ_e^l are respectively the thresholds of binary classifiers that targets to detect starts and ends of triggers' labels. If $P_{Ts}^l(c_i) > \delta_s^l$, token c_i is identified as the start of type label l. If $P_{Te}^l(c_i) > \delta_e^l$, token c_i is identified as end of type label l.

3.2 Argument Extraction

Once the triggers and their type labels have been identified, we come to the argument extraction component. Argument classification is converted to event relation extraction for triple $<t, r, a>$. Note that when the sentence is identified as "Other" type, we simply skip the following operation for argument role extraction. To better learn the inter-dependencies among the multiple events appearing in one sentence, we randomly pick one of predicted triggers in a sentence during the training phase, while in the evaluation phase, all the predicted triggers are picked in turn to predict corresponding arguments and roles played in the triggering events. We integrate information of predicted trigger word to argument extraction component. In ACE corpus, more than 98.5% triggers contain no more than 3 characters, so we simply pick the embedding vectors of start $\beta_s(c_i)$ and end $\beta_e(c_j)$ of one predicted trigger word t, and then generate representation of trigger word $\beta(t)$ by averaging these two vectors.

$$\beta(t) = \frac{(\beta_s(c_i) + \beta_e(c_j))}{2} \tag{4}$$

When obtain representations of trigger words $\beta(t)$, we add original embedding generated by BERT and $\beta(t)$ together:

$$\beta^{'}(s) = \beta(s) + \beta(t) \tag{5}$$

After integrate information of predicted trigger word to BERT sentence encoding, feed $\beta^{'}(s)$ into a full-connected layer with sigmoid activation function. A token c_k is predicted as the start of an argument triggered by word t which plays role r with probability:

$$P_{As}(c_k, r|t) = \sigma(W_{As}^r \beta^{'}(c_k) + b_{As}^r) \tag{6}$$

while as the end triggered by word t with probability:

$$P_{Ae}(c_k, r|t) = \sigma(W_{Ae}^r \beta^{'}(c_k) + b_{Ae}^r) \tag{7}$$

where W_{Ae} and b_{As} are respectively the trainable weights and bias of binary classifier that targets to detect starts of arguments' roles, while W_{Ae} and b_{Ae} are respectively the trainable weights of the other binary classifier that detects ends of arguments' roles. Set thresholds of detecting starts and ends as $\varepsilon^r = \{\varepsilon_s^r, \varepsilon_e^r\}$, ε_s^r and ε_e^r are respectively the thresholds of binary classifiers that target to detect starts and ends of triggers' labels. If $P_{As}(c_k, r|t) > \varepsilon_s^r$,token c_k is identified as the start of argument role r. If $P_{Ae}(c_k, r|t) > \varepsilon_e^r$, token c_k is identified as the end of argument role r.

3.3 Model Training

We train the joint model and define L_T as the loss function of all binary classifiers that are responsible for detecting triggers, shown as follows:

$$L_T = \frac{1}{m \times n}(\sum_{l=0}^{m}\sum_{i=0}^{n} -logP_{Ts}^l(c_i) + \sum_{l=0}^{m}\sum_{i=0}^{n} -logP_{Te}^l(c_i)) \tag{8}$$

L_T denotes the average of cross entropy of output probabilities of all binary classifiers which detect starts and ends of triggers on each type label. In the same way, we define L_A as the loss function of all binary classifiers that are responsible for detecting event relation triples:

$$L_A = \frac{1}{m \times n}(\sum_{r=0}^{m}\sum_{i=0}^{n} -logP_{As}(c_k, r|t) + \sum_{r=0}^{m}\sum_{i=0}^{n} -logP_{Ae}(c_k, r|t)) \tag{9}$$

where m denotes the sum of event label types and argument role types. L_A denotes the average of cross entropy of output probabilities of all binary classifiers which detect starts and ends of arguments on each role. The final loss function $L_E = L_T + L_A$. We minimize the final loss function to optimize the parameters of the model.

4 Experiments

We evaluate JMCEE framework on the ACE 2005 dataset that contains 633 Chinese documents. We follow the same setup as [1,2,13], in which 549/20/64 documents are used for training/development/test set. The proposed model is compared with the following state-of-the-art methods:

1) DMCNN [12] adopts dynamic multi-pooling CNN to extract sentence-level features automatically.
2) Rich-C [9] is a joint-learning, knowledge-rich approach including character-based features and discourse consistency features.
3) C-BiLSTM [13] designs a convolutional Bi-LSTM model which conduct Chinese event extraction from perspective of a character-level sequential labeling paradigm.
4) NPNs [1] performs event extraction in a character-wise paradigm, where a hybrid representation is learned to capture both structural and semantic information from both characters and words.

ACE 2005 dataset annotates 33 event subtypes and 35 role classes. The tasks of event trigger classification and argument classification in this paper are combined into a 70-category task along with "None" word and "Other" type. In order to evaluate the effectiveness of our proposed model, we evaluate models by micro-averaged Precision (P), Recall (R) and F1-score followed the computation measures of Chen and Ji (2009). It is worth noting that all the predicted roles for an argument are required to match with the golden labels, instead of just one of them. We take a further step to see the impacts of pipelined model and joint model. The pipelined model called MCEE which identifies triggers and arguments in two separate stages based our classification algorithm. The highest F-score parameters on the development set are picked and listed in Table 1.

Table 2 shows the results of trigger extraction on ACE 2005. The performances of Rich-C and C-BiLSTM are reported in their papers. As is seen, our JMCEE framework achieves the best F1 scores for trigger classi cation among

Table 1. Hyper-parameters for experiments.

Hyper-parameter	Trigger classification	Argument classification
Character embedding	768	768
Maximum length	510	510
Batch size	8	8
Learning rate of Adam	0.0005	0.0005
Classification thresholds	[0.5, 0.5, 0.5, 0.5]	[0.5, 0.4, 0.5, 0.4]

all the compared methods. Our JMCEE gains at least 8% F1-score improvements on trigger classification task on ACE 2005, which steadily outperforms all baselines. The improvement on the trigger extraction is quite significant, with a sharp increase of near 10% on the F1 score compared with these conventional methods. Table 3 shows results of argument extraction. Compared with these baselines, our JMCEE is at least 3% higher over other models on F1-score on argument classification task. While the improvement in argument extraction is not so obvious comparing to trigger extraction. This is probably due to the rigorous evaluation metric we have taken and the difficulty of argument extraction. Note that by our approach we identify 89% overlap roles in test set. Moreover, results show that our joint model substantially outperforms the pipelined model whether on trigger classification or argument classification. It is seen that joint model enables to capture the dependencies and interactions between the two subtasks and communicate deeper information between them, and thus improves the overall performance.

Table 2. Comparison of different methods on Chinese trigger extraction on ACE 2005 test set. Bold denotes the best result.

Model	Trigger identification			Trigger classification		
	P	R	F1	P	R	F1
DMCNN	66.6	63.6	65.1	61.6	58.8	60.2
Rich-C	62.2	71.9	66.7	58.9	68.1	63.2
C-BiLSTM	65.6	66.7	66.1	60.0	60.9	60.4
NPNs	75.9	61.2	67.8	73.8	59.6	65.9
MCEE (BERT-Pipeline)	82.5	78.0	80.2	72.6	68.2	70.3
JMCEE (BERT-Joint)	**84.3**	**80.4**	**82.3**	**76.4**	**71.7**	**74.0**

Table 3. Comparison of different methods on Chinese argument extraction on ACE 2005 test set. Bold denotes the best result.

Model	Argument identification			Argument classification		
	P	R	F1	P	R	F1
Rich-C	43.6	**57.3**	49.5	39.2	**51.6**	44.6
C-BiLSTM	53.0	52.2	52.6	47.3	46.6	46.9
MCEE (BERT-Pipeline)	59.5	40.4	48.1	51.9	37.5	43.6
JMCEE (BERT-Joint)	**66.3**	45.2	**53.7**	**53.7**	46.7	**50.0**

5 Conclusions

In this paper, we propose a simple yet effective joint Chinese multiple events extraction framework which jointly extracts triggers and arguments. Our contribution in this work is as follows:

1) Event relation triple is defined and incorporated into our framework to learn inter-dependencies among event triggers, arguments and arguments roles, which solves the roles overlap problem.
2) Our framework performs event extraction in a character-wise paradigm by utilizing multiple sets of binary classifiers to determine the spans, which allows to extract multiple events and relation triples and avoids Chinese language specific issues.

Our future work will focus on data generation to enrich training data and try to extend our framework to the open domain.

References

1. Chen, Z., Ji, H.: Language specific issue and feature exploration in Chinese event extraction. In: Proceedings of NAACL-HLT 2009, pp. 209–212 (2009)
2. Lin, H.Y., et al.: Nugget proposal networks for Chinese event detection. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 1565–1574 (2018)
3. Yang, S., et al.: Exploring pre-trained language models for event extraction and generation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 5284–5294 (2019)
4. Doddington, G.R., et al.: The automatic content extraction (ACE) program-tasks, data, and evaluation. In: LREC (2004)
5. He, R.F., Duan, S.Y.: Joint Chinese event extraction based multi-task learning. J. Softw. **30**(4), 1015–1030 (2019)
6. Liu, X., Luo, Z., Huang, H.: Jointly multiple events extraction via attention-based graph information aggregation. arXiv preprint arXiv:1809.09078 (2018)
7. Nguyen, T.H., Cho, K., Grishman, R.: Joint event extraction via recurrent neural networks. In: Proceedings of NAACL-HLT 2016, pp. 300–309 (2016)

8. Nguyen, T.M., Nguyen, T.H.: One for all: neural joint modeling of entities and events. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 6851–6858 (2019)
9. Chen, C., Ng, V.: Joint modeling for Chinese event extraction with rich linguistic features. In: Proceedings of COLING 2012, pp. 529–544 (2012)
10. Liao, S., Grishman, R.: Using document level cross-event inference to improve event extraction. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 789–797 (2010)
11. Li, Qi., Ji, H., Huang, L.: Joint event extraction via structured prediction with global features. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 73–82 (2013)
12. Chen, Y.B., et al.: Event extraction via dynamic multi-pooling convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, vol.1, pp. 167–176 (2015)
13. Zeng, Y., Yang, H., Feng, Y., Wang, Z., Zhao, D.: A convolution BiLSTM neural network model for Chinese event extraction. In: Lin, C.-Y., Xue, N., Zhao, D., Huang, X., Feng, Y. (eds.) ICCPOL/NLPCC -2016. LNCS (LNAI), vol. 10102, pp. 275–287. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-50496-4_23
14. McCann, B., et al.: Learned in translation: contextualized word vectors. In Advances in Neural Information Processing Systems, pp. 6294–6305 (2017)
15. Peters, M., et al.: Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, vol. 1, pp. 2227–2237 (2018)
16. Radford, A., et al.: Improving language understanding by generative pre-training (2018). https://www.cs.ubc.ca/amuham01/LING530/papers/radford2018improving.pdf
17. Devlin, J., et al.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
18. Sha, L., et al.: Jointly extracting event triggers and arguments by dependency-bridge RNN and tensor-based argument interaction. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, pp. 5916–5923 (2018)